Quiz 1 - Detailed Solutions

Question 1: Vector Orthogonality and Similarity

Given: $u = [2, -3]^T$ and $v = [6, 4]^T$

Part A: Are these vectors orthogonal?

Definition: Two vectors are orthogonal (perpendicular) if and only if their dot product equals zero.

Step 1: Calculate the dot product $u \cdot v$

$u \cdot v = u_1 v_1 + u_2 v_2$
$u \cdot v = (2)(6) + (-3)(4)$
$u \cdot v = 12 + (-12)$
$\mathbf{u \cdot v = 0}$

Step 2: Interpret the result

Since $u \cdot v = 0$, the vectors are orthogonal.

Answer: YES, the vectors u and v are orthogonal because their dot product equals zero.

Part B: Are these vectors similar?

Definition: In this context, "similar" vectors typically means parallel vectors (one is a scalar multiple of the other): $v = k \cdot u$ for some scalar k.

Step 3: Check if $v = k \cdot u$ for some constant k

If $v = k \cdot u$, then each component must satisfy:
$v_1 = k \cdot u_1 \rightarrow 6 = k \cdot 2 \rightarrow k = 3$
$v_2 = k \cdot u_2 \rightarrow 4 = k \cdot (-3) \rightarrow k = -4/3$

Step 4: Compare the k values

We get $k = 3$ from the first component and $k = -4/3$ from the second. These are not equal.

Answer: NO, the vectors are not similar (not parallel) because there is no single scalar k that satisfies $v = k \cdot u$. In fact, since they are orthogonal, they point in perpendicular directions.

Part C: Write a line expression to separate these points

Interpreting the vectors as points: $P_1 = (2, -3)$ and $P_2 = (6, 4)$

$w^T x + b = c$

Three valid approaches exist, each optimizing a different objective:

Method 1: Simple Line Through Origin

Step 5: Find direction vector: $d = P_2 - P_1 = (4, 7)$
Step 6: Use as normal vector with c = 0: $4x + 7y = 0$

$(w_1, w_2)x + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = 0$

Step 7: Verify: $P_1 \rightarrow 4(2) + 7(-3) = -13$ (negative), $P_2 \rightarrow 4(6) + 7(4) = +52$ (positive)

Answer: $4x + 7y = 0$ (simplest form, passes through origin)

Method 2: Perpendicular Bisector (Maximum Margin)

Step 8: Find midpoint: $M = ((2+6)/2, (-3+4)/2) = (4, 0.5)$
Step 9: Normal vector: $n = (4, 7)$, Line passes through M
Step 10: Compute c: $c = 4(4) + 7(0.5) = 16 + 3.5 = 19.5$
Step 11: Verify equal distance: $P_1 \rightarrow -13 - 19.5 = -32.5$, $P_2 \rightarrow 52 - 19.5 = +32.5$

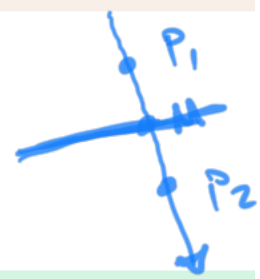
Answer: $4x + 7y = 19.5$ (equidistant from both points, like SVM)

Method 3: Pseudo-Inverse (Least Squares Optimal)

X^+

Step 12: Set up system $Xw = y$ with targets $y = [-1, +1]^T$
$X = [[2, -3], [6, 4]]$ (each row is a point)
Step 13: Solve using pseudo-inverse: $w = X^+y = \boxed{(X^TX)^{-1}X^Ty}$
$w = [-0.0385, 0.3077]^T$ or equivalently $w = [-1, 8]^T$ (scaled)
Step 14: Verify: $w \cdot P_1 = -1$ (exact!), $w \cdot P_2 = +1$ (exact!)

Answer: $\boxed{-x + 8y = 0}$ (least squares fit to targets [-1, +1])



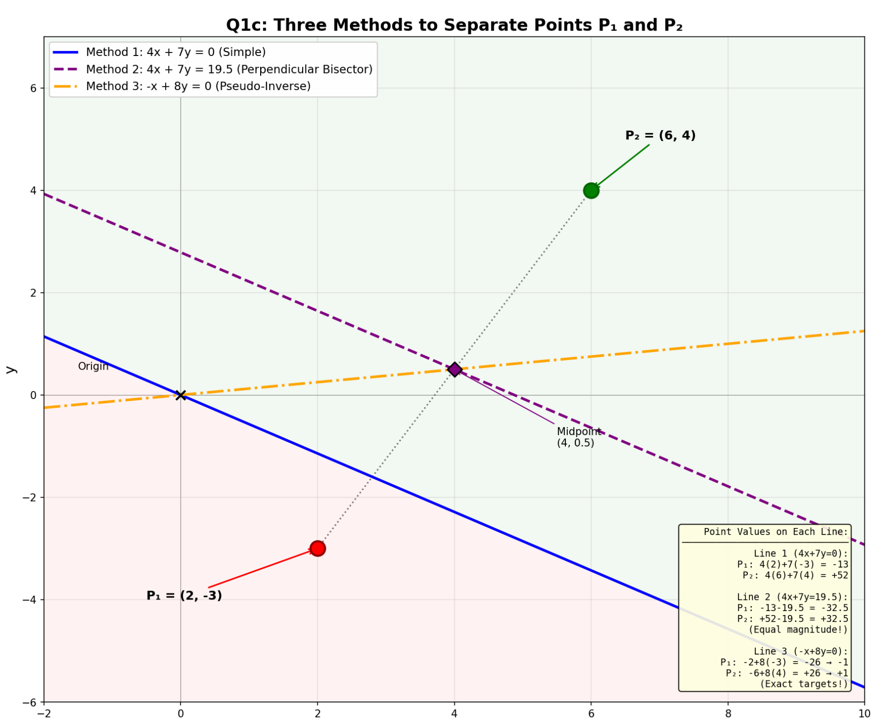
$P_1 - P_2$

$w^T x + y$ =

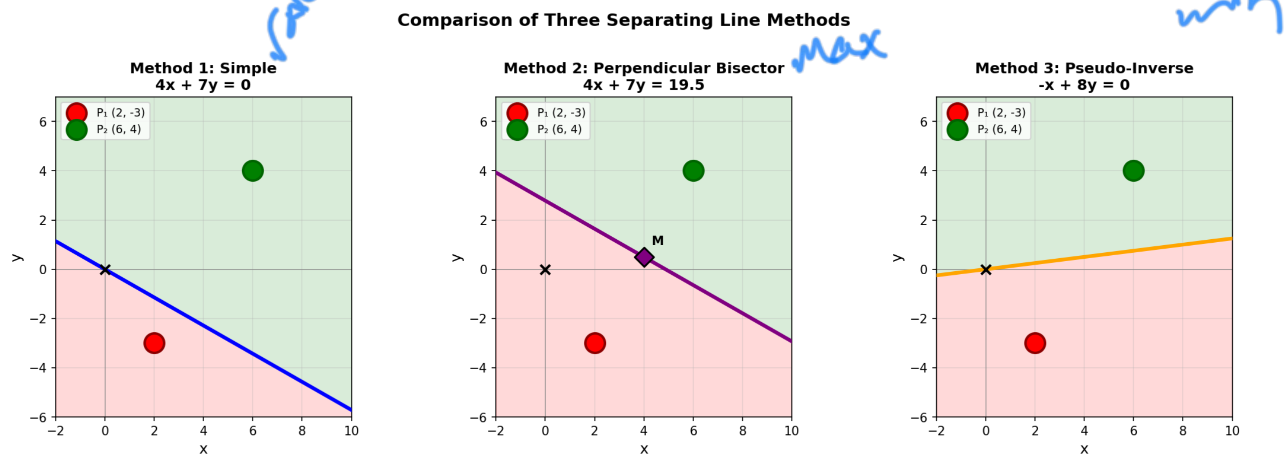
Summary - All Valid Answers:

- **$4x + 7y = 0$** — Simple (through origin)
- **$4x + 7y = 19.5$** — Perpendicular bisector (maximum margin)
- **$-x + 8y = 0$** — Pseudo-inverse (least squares optimal)

Note: Any line $4x + 7y = c$ with $-13 < c < 52$ separates the points. Different methods optimize different objectives.



random max with



Question 2: Decision Boundary Parameter

Given: Linear model $y = 2x_0 - x_1 + b$. Find b such that the decision boundary passes through points (1, 4) and (3, 2).

Key Concept: The decision boundary is where $y = 0$, giving us: $2x_0 - x_1 + b = 0$

Step 15: Apply the decision boundary equation to point (1, 4) where $x_0 = 1$, $x_1 = 4$

$2(1) - 4 + b = 0$
$2 - 4 + b = 0$
$\mathbf{b = 2}$

Step 16: Apply the decision boundary equation to point (3, 2) where $x_0 = 3$, $x_1 = 2$

$2(3) - 2 + b = 0$
$6 - 2 + b = 0$
$\mathbf{b = -4}$

Step 17: Analyze the results

We obtain b = 2 from the first point and b = -4 from the second point.

Important Observation: These constraints are inconsistent. The two points do not lie on any line of the form $2x_0 - x_1 + b = 0$ (a line with fixed normal vector [2, -1]).

Answer: No single value of b exists that makes the decision boundary pass through both points. If the question expects a unique answer, there may be a typo. If forced to choose: b = 2 for point (1,4) or b = -4 for point (3,2).

Question 3: Supervised Learning True/False

Statement: "In supervised machine learning, we need training data with known outputs (ground truth) to estimate model parameters."

Step 18: Define supervised learning

Supervised learning is a machine learning paradigm where the model learns from labeled examples, i.e., input-output pairs (x_i, y_i).

Step 19: Explain why ground truth is essential

- The model makes predictions $\hat{y} = f(x; \theta)$ using parameters θ
- We compute the loss $L(y, \hat{y})$ comparing predictions to ground truth
- Parameters are updated to minimize this loss via gradient descent
- Without known outputs, we cannot compute loss or update parameters

Answer: TRUE. Supervised learning requires labeled training data because the learning algorithm needs ground truth outputs to compute prediction errors and optimize model parameters through loss minimization.

Question 4: Line Equation in Vector Form

Given: The equation $3x + 4y = 12$ represents a line in 2D.

w

$w^T x = 0$

Part A: Write in vector form $n^T x = d$

Step 20: Identify the components

The equation $ax + by = c$ can be written as $n^T x = d$ where:

$n = [a, b]^T$ (normal vector, coefficients of x and y)

$x = [x, y]^T$ (position vector)

$d = c$ (constant on right side)

Step 21: Apply to our equation $3x + 4y = 12$

$\mathbf{n = [3, 4]^T}$ **and** $\mathbf{d = 12}$

Answer: $[3, 4]^T \cdot [x, y]^T = 12$, where the normal vector $n = [3, 4]^T$

Part B: What is the unit normal vector $\hat{n}$?

Definition: The unit normal vector has magnitude 1 and points in the same direction: $\hat{n} = n / ||n||$

Step 22: Calculate the magnitude $||n||$

$||n|| = \sqrt{(n_1^2 + n_2^2)}$

$||n|| = \sqrt{(3^2 + 4^2)}$

$||n|| = \sqrt{(9 + 16)}$

$||n|| = \sqrt{25} = 5$

$\vec{v}$

$\hat{v}$

$\sqrt{0.6^2 + 0.8^2} = 1$

Step 23: Compute the unit normal

$\hat{n} = n / ||n|| = [3, 4]^T / 5$

$\mathbf{\hat{n} = [3/5, 4/5]^T = [0.6, 0.8]^T}$

Answer: $\mathbf{\hat{n} = [3/5, 4/5]^T}$ or equivalently $[0.6, 0.8]^T$. This is the famous 3-4-5 right triangle!

Question 5: Deep Neural Network Geometry

Question: Which best describes what a deep neural network does geometrically?

Options: (1) Single linear transformation, (2) Multiple nonlinear transformations, (3) Memorizes training data, (4) Creates lookup table

Step 24: Analyze each option

- **Option 1:** False. A single linear transformation can only create linear decision boundaries. Multiple linear transformations without nonlinearity still produce a single linear transformation (matrix multiplication is associative).
- **Option 2:** Correct. Each layer applies a linear transformation followed by a nonlinear activation (ReLU, sigmoid, etc.). This allows the network to learn complex, curved decision boundaries.
- **Option 3:** False. While overfitting can lead to memorization, the goal is generalization. A properly trained network learns patterns, not memorizes examples.
- **Option 4:** False. A lookup table would require storing every possible input. Neural networks generalize to unseen inputs through learned representations.

Step 25: Geometric interpretation

Each layer transforms the feature space. The nonlinear activations allow the network to "bend" and "fold" the space, making classes that were previously intertwined become linearly separable in the transformed space.

Answer: (2) Applies multiple nonlinear transformations to map input points to an output space where classification is easier. This is the fundamental geometric insight behind deep learning.

Question 6: Eigenvalue and Eigenvector Computation

Given: $\lambda = 3$ is ~~an eigenvalue of matrix B with eigenvector~~ $e = [1, 2]^T$. Compute Be.

Step 26: Recall the eigenvalue-eigenvector definition

$Bv = \lambda v$ where λ is an eigenvalue and v is the corresponding eigenvector.

This equation states that multiplying an eigenvector by its matrix produces a scalar multiple of that eigenvector.

Step 27: Apply the definition to our problem

$Be = \lambda e$

$Be = 3 \cdot [1, 2]^T$

$Be = [3, 6]^T$

Step 28: Verify understanding

Notice that Be points in the same direction as e (both along the line $y = 2x$), just scaled by factor 3. This is the geometric meaning of eigenvectors: they are directions that remain unchanged under the transformation, only stretched or compressed.

Answer: $Be = [3, 6]^T$

$\underline{B\vec{e}} = \lambda\vec{e}$

Summary of Answers

Question	Answer
Q1a	YES - vectors are orthogonal (dot product = 0)
Q1b	NO - vectors are not similar/parallel (no common scalar)
Q1c	4x+7y=0 (simple) | 4x+7y=19.5 (bisector) | -x+8y=0 (pseudo-inv)
Q2	No unique b exists (b=2 for (1,4), b=-4 for (3,2))
Q3	TRUE - supervised learning requires labeled data
Q4a	$n = [3, 4]^T$, $d = 12$
Q4b	$\hat{n} = [3/5, 4/5]^T = [0.6, 0.8]^T$
Q5	Option (2) - multiple nonlinear transformations
Q6	$Be = [3, 6]^T$